Biochemistry (revised)

A Novel Activity of *Escherichia coli* Mismatch Uracil-DNA Glycosylase (Mug) Excising 8-(Hydroxymethyl)-3,N^4-ethenocytosine, a Potential Product Resulting from Glycidaldehyde Reaction†

Bo Hang*, Gary Downing, Anton B. Guliaev and B. Singer

Donner Laboratory, Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720

Running Title: 8-HM-εC: a New Substrate for *E. coli* Mug Protein

† This work was supported by NIH grant CA72079 (to B.H.) and was administered by the Lawrence Berkeley National Laboratory under Department of Energy contract DE-AC03-76SF00098.

*To whom correspondence should be addressed.

Telephone: (510) 486-7268
Fax: (510) 486-6488
E-mail: Bo_Hang@lbl.gov

Abbreviations: 8-HM-εC, 8-(hydroxymethyl)-3,N^4-ethenocytosine; εC, 3,N^4-ethenocytosine; AP, apurinic/apyrimidinic; Mug, mismatch uracil-DNA glycosylase; TDG, mismatch specific thymine-DNA glycosylase; Ung, uracil-DNA glycosylase; APNG, alkylpurine-DNA-N-glycosylase; PAGE, polyacrylamide gel electrophoresis; DMT, 4,4´dimethoxytritylchloride; TEAA, triethylammoniumacetate.

ABSTRACT: Glycidaldehyde is an industrial chemical which has been shown to be genotoxic in *in vitro* experiments and carcinogenic in rodent studies. It is a bifunctional alkylating agent capable of reacting with DNA to form exocyclic hydroxymethyl-substituted ethenobases. In this work, 8-(hydroxymethyl)-3,N^4-etheno-2'-deoxycytidine (8-HM-εdC), a potential nucleoside derivative derived from glycidaldehyde, was synthesized using phosphoramidite chemistry and site-specifically incorporated into a defined 25-mer oligodeoxynucleotide. The 8-HM-εC adduct is structurally related to 3,N^4-ethenocytosine (εC), a product of reaction with vinyl chloride or through lipid peroxidation. In *Escherichia coli*, εC has been shown previously to be a primary substrate for the mismatch uracil-DNA glycosylase (Mug). In this study we report that the same glycosylase also acts on 8-HM-εC in an oligonucleotide duplex. The enzyme binds to the 8-HM-εC-oligonucleotide to a similar extent as the εC-oligonucleotide. The Mug excision activity toward 8-HM-εC is ~2.5-fold lower than that toward the εC substrate. Both activities can be stimulated to ~2-fold higher by the addition of *E. coli* endonuclease IV. These two adducts, when mispaired with normal bases, were all excised from DNA by Mug with similar efficiencies. Structural studies using molecular simulations showed similar adjustment and hydrogen bonding pattern for both 8-HM-εC•G and εC•G pairs in oligomer duplexes. We believe that these findings may have biological and structural implications in defining the role of 8-HM-εC in glycosylase recognition/repair.

INTRODUCTION

Simple epoxides represent an important group of industrial chemicals. Some of these compounds are mutagenic and/or carcinogenic. For example, glycidaldehyde, a highly reactive epoxide (Figure 1), has been shown to be both mutagenic in *in vitro* genotoxicity tests (1) and carcinogenic in long-term rodent skin cancer studies (2,3). Based on these studies and others, the International Agency for Research on Cancer (IARC) classified glycidaldehyde as an animal carcinogen (4,5). Studies from National Toxicology Program (NTP) also showed sufficient evidence for the carcinogenicity of glycidol in experimental animals and anticipated that these biological effects could occur in humans exposed to the compound (6). Glycidaldehyde can be produced from oxidative metabolism of glycidyl ethers (7,8). Several glycidyl ethers have also been shown to be carcinogenic in experimental animals (9,10).

Due to its reactive carbonyl and epoxy functionalities, glycidaldehyde is capable of forming cyclic hydroxymethyl-substituted etheno adducts. The structures of the dA and dG adducts formed after reaction with glycidaldehyde have previously been identified and well characterized (11-16). The modified dA nucleoside was identified by Steiner and colleagues in the skin of C3H mice treated with glycidaldehyde (13) or bisphenol A diglycidyl ether (8). The only report on the chemical formation of dC-glycidaldehyde adducts is from Kohwi (17), who observed that glycidaldehyde is highly reactive with dC in non-B DNA *in vitro*. However, no structural information was given on the dC modification.

In order to understand the role of a dC adduct in the mechanism of mutagenicity/ carcinogenicity of glycidaldehyde as well as its chemical/physical properties, we recently synthesized 8-(hydroxymethyl)-3,N^4-etheno-2'-deoxycytidine (8-HM-εdC) (Figure 1) and its phosphoramidite (18). The latter compound was then site-specifically incorporated into

defined oligodeoxynucleotides. Using an *in vitro* replication assay, we have found that the presence of 8-HM-εC either causes blocking of replication or facilitates translesional syntheses catalyzed by mammalian DNA polymerases, mainly in an error-prone manner (19). Some of these polymerases such as pol η were able to catalyze significant amounts of lesion-bypass with miscoding opposite the adduct. This strongly suggests that 8-HM-εC could be a promutagenic lesion *in vivo*, if such adducts are indeed formed in genomic DNA and not repaired.

8-HM-εC is a structural analog of the exocyclic adduct 3,N^4-etheno C (εC), in which the 5-membered etheno ring has a -CH_2OH group substituting a hydrogen (Figure 1). εC is mainly formed from environmental compounds such as vinyl chloride and ethyl carbamate or through the process of lipid peroxidation (for review, see ref 20). Both in *E. coli* and mammalian systems, the εC adduct in DNA has been shown to be miscoding *in vitro* and promutagenic *in vivo* (for review, see ref 21). In our recent study using mammalian DNA polymerases (19), εC showed similar miscoding specificity to that of 8-HM-εC.

In 1994, this laboratory first reported a DNA glycosylase activity in HeLa cell-free extracts which released εC as a free base (22). During further purification of this activity, we found that it actually resides in a protein different from the glycosylase acting on 1,N^6-ethenoadenine, namely alkylpurine-DNA-N-glycosylase (APNG) (23). This finding was further confirmed by the observation that the εC-DNA glycosylase activity was unchanged in mouse tissues regardless of whether or not the APNG gene was knocked out (24). The protein excising εC has now been identified as the mismatch-specific thymine-DNA glycosylase (TDG) in humans (25,26) and mismatch uracil-DNA glycosylase (Mug) in *E. coli* (26), which is the homologue of the human TDG (27). One difference between these two enzymes is that Mug lacks the G•T mismatch repair activity (27). The εC activity was also observed in a recombinant G•T mismatch DNA glycosylase from the thermophilic

bacterium, *Methanobacterium thermoautotrophicum* THF (25), a functional homologue of TDG/Mug proteins (28). These enzymes were originally found to remove uracil and/or thymine from duplexes when paired with guanine (28-31). Both TDG and Mug are capable of excising εC from oligomer duplexes efficiently (25,26,32), with Mug having greater efficiency toward the adduct (26).

In this work, we report that 8-HM-εC is also a good substrate for the *E. coli* Mug protein. For comparative purposes, the previously identified substrate for the enzyme, εC, was studied in parallel using *in vitro* assays. In addition, molecular modeling of these two closely related exocyclic derivatives in DNA duplexes was performed in order to obtain structural data which may aid in our understanding of enzymatic recognition of the two adducts.

EXPERIMENTAL PROCEDURES

Materials

Crotalus adamanteus venom phosphodiesterase 1, bacterial alkaline phosphatase, and [γ-^{32}P] ATP (specific activity 6,000 Ci/mmol; 1 Ci = 37 GBq) were purchased from Amersham Pharmacia Biotech. T4 polynucleotide kinase was purchased from United States Biochemical. Acetonitrile HPLC grade, and glacial acetic acid were obtained from Fisher Scientific. Triethylamine was purchased from Aldrich. Sep-Pak C18 Cartridges were purchased from Waters. The unmodified phosphoramidites for ultramild deprotection: 4-isopropyl-phenoxyacetyl (iPr-Pac)–dG, phenoxyacetyl(Pac)–dA, acetyl dC, and dT-CE phosphoramidite, and the dT ABI Controlled Pore Glass (CPG) 500Å column were purchased from Glen Research.

Repair Enzymes and Cell-free Extracts

E. coli Mug protein, MutY protein, formamidopyrimidine-DNA glycosylase (Fpg protein), endonuclease III (EndoIII), endonuclease VIII (EndoVIII) were from Trevigen (Gaithersburg, MD). One unit of Mug protein is defined as the amount of enzyme required to cleave 1 pmole of a ^{32}P-labeled oligomer containing an εC in a duplex in 1 hr at 37°C. *E. coli* uracil-DNA glycosylase (Ung) was from Amersham Pharmacia Biotech. The major human AP endonuclease 1 (HAP1) was a gift from Dr. I. D. Hickson (University of Oxford). *E. coli* endonuclease IV (EndoIV) was a gift from Dr. D. M. Wilson (Lawrence Livermore Laboratory). *E. coli* strain BW32 (AB1157), used to prepare a wild-type cell-free extract, was a gift from Dr. B. Weiss (University of Michigan). The preparation of crude extract was carried out as previously described by Hang *et al* (33).

8-HM-εdC and Oligonucleotide Synthesis

The preparation of 8-(Hydroxymethyl)-3,N^4-etheno-2'-deoxycytidine (8-HM-εdC) and its phosphoramidite was performed using a modification of the procedures described by Chenna *et al* (18). The 25-mer oligodeoxynucleotide containing 3,N^4-ethenocytosine (εC) was synthesized as described previously (34). Unmodified oligodeoxynucleotides were synthesized and purified by Operon, Inc. (Alameda, CA).

The synthesis of the 25-mer 5´-CCGCTAGXGGGTACCGAGCTCGAAT-3´ (X = 8-HM-εC) was carried out on an Applied Biosystems 392 automated DNA synthesizer on a 1 μmol scale, ABI column, using phosphoramidites which can be deprotected under ultra mild conditions. The coupling time for the modified nucleoside was increased to 900s to give an optimal coupling efficiency of 93%. The 5´-DMT-on 25-mer oligodeoxynucleotide was

base-deprotected and cleaved from the resin, under strictly anhydrous conditions, and in the absence of light, using 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) dried over 3Å molecular sieves with anhydrous methanol (100 μl of DBU to 900 μl of dry methanol). The deprotection was complete after 12 hr stirring at room temperature.

Purification of the DMT-on oligomer was carried out on a Luna 5μ Phenyl Hexyl 250 x 4.60 mm analytical column (Phenomenex). The DMT-on oligomer was purified using a system in which the acetonitrile concentration was maintained at 15% for 10 min in the presence of triethylammonium acetate (TEAA) buffer, then increased linearly to 35% over the next 25 min. The flow rate was maintained at 1 ml/min. Using this system the pure DMT-on oligomer had a retention time of 26.82 min.

The isolated DMT-on oligomer was treated with 80% acetic acid for 10 min, neutralized, evaporated to dryness, and purified again on the same column with the DMT-off. The resultant peak was collected using a second system where the acetonitrile was maintained at 5% in TEAA for 10 min then increased linearly to 30% acetonitrile after 45 min. The retention time for the DMT-off oligodeoxynucleotide was 24.66 min. Analysis of this compound by electrospray mass spectrometry run on a VG Bio-Q Instruments mass spectrometer found the correct peak m/z -7722.51 M^+.

Further verification of the incorporation of the correct modified base was performed by enzyme digestion of the above 25-mer according to the procedure defined in ref 35. The reaction mixture was then analyzed by HPLC (Hewlett Packard 1100 detector) using a reverse-phase Luna 5 μ Phenyl Hexyl 250 x 4.60 mm analytical column. The result showed that by “spiking” the sample mixture with the modified monomer an enlarged peak was observed superimposed over the modified monomer peak from the digested oligonucleotide at 19.39 min, confirming that the synthesized adduct was incorporated.

Band Shift and Glycosylase Assay

For testing protein binding and enzymatic activities, both modified and unmodified 25-mer oligonucleotides were 5'-end labeled with [γ-^{32}P] ATP and annealed to a complementary oligonucleotide in a 1 to 1.5 ratio as previously described (36). The standard binding reactions contained 1.5 nM 5'-end ^{32}P-labeled duplex in 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM EGTA, 0.1 mg/ml acetylated BSA and varying concentrations of Mug protein in a total volume of 10 µl. After 15 min incubation at room temperature (~20 °C), reactions (5 µl) were resolved on 6% nondenaturing polyacrylamide gel electrophoresis (PAGE) for 1 hr at 150 V using 1X TBE buffer.

To measure Mug excision activity, the same reaction mixtures as described above for the binding assay were incubated at 37 °C with varying amounts of enzyme or for various lengths of time. The reactions were stopped by heating the samples at 95-100 °C for 3 min and then placing on ice. Following this, a 5' AP endonuclease, HAP1 or EndoIV, was added to the reactions, which were incubated at 37 °C for a further 20 min. This step was used to cleave the apyrimidinic (AP) site resulting from the excision of 8-HM-εC or εC by Mug. HAP1/EndoIV alone did not exhibit any detectable effect on these two substrates. In the reactions with AP lyase-containing glycosylases including Fpg protein, EndoIII and EndoVIII, there was no addition of HAP1. One exception is MutY protein among the *E. coli* glycosylases tested. Whether this enzyme has an associated AP lyase activity has been the subject of controversy (*e.g*. 37-40). Therefore, HAP1 was added to the MutY reaction. When excision reactions were performed with cell-free extracts of *E. coli* AB1157, no AP endonuclease was added. For the competition assay, the same unlabeled 25-mer oligomers were pre-incubated with Mug protein in the mixture for 5 min on ice before ^{32}P-labeled 8-HM-εC-containing oligomer duplex was added. All the above reactions were stopped by

adding equal amounts of a F/E solution (90% formamide plus 50 mM EDTA) and then heated at 95-100 °C for 3 min. Samples (5 µl/lane) were then run on 12% 8 M urea/PAGE and subsequently the gel was dried and autoradiographed. For band quantitation, the Bio-Rad FX Molecular PhosphorImager and Quantity One software (version 4.0.1) were used.

Molecular Modeling

The modified bases (8-HM-εC and εC) were constructed using the xLeap module of AMBER 5.0 (41). A set of parameters for the modified bases, including atom charges, bond and torsional angles, and bond stretching constants, have been developed based on *ab initio* quantum mechanical calculations, as described by Guliaev *et al* (42). The modified bases were incorporated into 15-mer DNA duplexes truncated from the 25-mer sequence (see 8-HM-εdC and Oligonucleotide Synthesis): 5'-CCGCTAG**X**GGGTACC-3', where X = 8-HM-εC, εC or C. The topology and coordinate files for all three duplexes were generated using the xLeap module of AMBER 5.0. 28 Na^+ ions were placed around phosphate groups to neutralize negative charges and a rectangular water box was added which provided at least 10Å of explicit water molecules around each of the DNA duplexes. The system was subjected to a series of equilibration runs to obtain the correct density and volume for the water present. Finally, 1ns of the unrestrained molecular dynamics runs, using particle mesh Ewald (PME) to treat electrostatic interactions, were used to generate averaged structures (42). The average structures from the last 650 ps of MD were analyzed using CURVES 5.3 (43) and MD trajectories were analyzed using the carnal module of AMBER 5.0. The structures were also displayed and analyzed using Insight II (Insight II 98.0, Biosym/MSI, San Diego, CA). All calculations were performed on Silicon Graphics Origin 200 server (Silicon Graphics Inc., Mountain View, CA).

RESULTS

Recognition and Excision of 8-HM-εC by E. coli Mug Protein

We started our initial binding experiments on 8-HM-εC using *E. coli* mismatch uracil-DNA glycosylase (Mug), since this is the only enzyme found so far to remove εC in *E. coli* (26,44). As shown in Figure 2, Mug protein showed a similar protein concentration-dependent binding activity toward a ^{32}P-end labeled 8-HM-εC-containing oligomer duplex (lanes 4-9) to that toward εC-containing duplex (lanes 11-16). In this 25-mer both 8-HM-εC and εC are in the same position (8th nucleotide from 5' end) and the opposite base is G (see EXPERIMENTAL PROCEDURES).

Mug protein was also shown to cleave the same 25-mer containing 8-HM-εC in a protein concentration-dependent manner but the extent of cleavage is lower than that of εC cleavage (Figure 3A). In these experiments, the cleavage products from both substrates after a 5' AP endonuclease treatment were a ^{32}P-labeled 7-mer, resulting from hydrolysis 5' to the AP site at the 8th position. Alkaline treatment (NaOH plus heat) of Mug-treated samples also caused incision of both 8-HM-εC and εC oligomer substrates. The cleavage pattern for the two adducts (data not shown) is the same which is also in agreement with that described for the β-elimination reaction that cleaves an AP site from the 3'-end (29,45). Both experiments indicate that Mug acts on 8-HM-εC as a DNA glycosylase.

The specificity of the 8-HM-εC activity of Mug was further confirmed using a competition assay. As shown in Figure 4, when the ^{32}P-end labeled 8-HM-εC 25-mer duplex was incubated with an increasing molar excess (1x, 2.5x, 5x and 10x) of the same but unlabeled competitor 25-mers, Mug activity toward 8-HM-εC could be efficiently competed

by the 8-HM-εC- or εC-containing duplex but not by the unmodified duplex. The cross-competition of 8-HM-εC excision by the εC oligomer confirmed that Mug is the enzyme acting on 8-HM-εC. Note that in Figure 4 the εC-containing 25-mer duplex (lanes 7 to 10) appears to be a better competitor than the 8-HM-εC-containing competitor (lanes 3 to 6), which can be correlated with the kinetic findings described below.

We further compared the kinetic features of Mug excision of 8-HM-εC with those of εC under the same assay conditions. Figure 3B shows the removal of 8-HM-εC (2 nM) and εC (2 nM) by Mug (0.5 nM) from the 25-mer duplexes as a function of time (0-60 min). Both activities had an early fast phase during the first 2.5 min and then reached a slower rate afterward. That rate of 8-HM-εC excision is slower than that of εC excision in both phases, with ~2.5-fold difference after 15 min reaction (Figure 3B). Addition of *E. coli* EndoIV, a 5' AP endonuclease next to the Mug protein in *E. coli* BER pathway, stimulated the 8-HM-εC excision by ~2-fold (Figure 3C). In both Figure 3B and C experiments, when 4 molar excess of substrate duplex was used, the 8-HM-εC or εC excised by Mug in the absence of EndoIV was less than the molar concentrations of Mug in these reactions. However, addition of 4 nM EndoIV to the reaction in Figure 3C which had 4 nM oligomer duplex and 1 nM Mug protein, more than 1 nM 8-HM-εC was excised, indicating the occurrence of turnover with the substrate.

In order to explore whether any other glycosylase in *E. coli* can act on 8-HM-εC, various available pure DNA glycosylases were tested. Except for Mug protein, no detectable activities were observed from other glycosylases tested (Figure 5). This test was carried out using high amounts of proteins compared to that required to cleave their normal substrates in order to detect weak activity. Figure 5 also shows that this newly identified 8-HM-εC activity resides in a wild-type *E. coli* strain. The same pattern was also observed for εC substrate specificity.

Effect of Opposite Bases and Strandness on Excision Efficiency

The adduct-containing 25-mers were 5'-end ^{32}P-labeled and annealed to a complementary strand with each of the four bases opposite 8-HM-εC or εC. Three *E. coli* DNA glycosylases were tested including Mug, Ung and MutY. Mug and MutY are known to recognize base mismatches. Table 1 summarizes the base pair specificity of the Mug protein as determined by the initial velocities of the adduct excision. All modified base mismatches were recognized and the adduct excised by Mug. For both adducts, the variation in activities toward these mismatches was within a relatively small range (Table 1). Unlike Mug, neither Ung nor MutY showed any detectable activity toward any of the modified base mismatches listed in Table 1, showing a clear difference in their substrate specificity from Mug.

With the single-stranded oligomer containing either 8-HM-εC or εC, no detectable cleavage was observed when up to 16 nM Mug protein was used for 1 h at 37^{o}C (data not shown). For these experiments, after Mug reaction with the adduct-containing oligonucleotide, a 2-fold molar excess of the complementary strand was added to the reaction which was then slowly cooled down from 80^{o}C to room temperature. HAP1 (5 ng) was then added and incubated at 37^{o}C to cleave any AP site in duplex DNA as this enzyme does not cleave AP site in single-stranded DNA (46).

Structural Features of 8-HM-εC and εC

The stability and the equilibrium state of the 1 ns MD simulations were evaluated by calculating the RMSD values of each 1ps 'snapshot' relative to the starting structures containing dC, 8-HM-εC or εC. All three structures reached the conformational equilibrium

in the range between 250 – 350 ps (data not shown). Therefore, the conformations generated between 350 and 1000 ps were used to monitor structural properties. All three structures remained in the B-DNA conformational family with the minor distortion around the lesion site during the entire course of the simulation. The dC adducts were displaced into the major groove of the helix, while the opposite base remained stacked.

The main feature of the two lesion-containing basepairs is the similar alignment of the bases in the pair with the high value for the shear (SHR) parameter (Table 2). The SHR values for the other basepairs in all three DNA duplexes, including control, fall into the range –0.2 to 0.3Å. The high SHR value correlated with the higher degree of opening of the lesion-containing basepair (Table 2). The alignment of the bases in the 8-HM-εC•G and εC•G pairs was stabilized by the observation of the single bifurcated hydrogen bond (yellow dotted lines in Figure 6). The stability of this bond was evaluated by calculating the percent of occupancy during the simulation. For the 8-HM-εC•G pair, the hydrogen bond between 8-HM-εC-O3 and G23-N2/N1 was 99.5% occupied with an average value of 3.98 ± 0.18 Å. In the case of εC•G basepair the hydrogen bond between εC-O2 and G23-N2/N1 was 95% occupied with an average value of 3.01 ± 0.2Å during the entire course of simulation. The observation of the highly sheared basepair for the εC containing duplexes was reported previously using NMR (47-50). These authors showed that εC opposite G, T or C formed a sheared base-pair with one hydrogen bond between the bases.

Molecular dynamics calculations also revealed a similar sugar conformation for the two adducts. These adducts, when present in the DNA duplex, have a sugar pucker in the C3'-endo/C4'-exo region, while the rest of the residues, including unmodified C, are in the C2'-endo/C3'-exo range (Figure 6). The C3'-endo/C4'-exo sugar conformation of the εC adduct produced by modeling was in a good agreement with the previously reported solution

structures of the εC containing DNA duplexes, which showed the same conformational range for that sugar (48,51).

DISCUSSION

Many chemicals can interact with cellular macromolecules and subsequently affect their functions. One of the most important factors in the initiation of carcinogenesis is the reaction of such chemicals with DNA bases to produce adducts or other types of damage (52). Glycidaldehyde, an animal carcinogen, has been shown to have the ability to modify bases both chemically (11-17) and in animal experiments (8,13). In this work, we studied the enzymatic repair of 8-(hydroxymethyl)-3,N^4-ethenocytosine (8-HM-εC) as it is a potential glycidaldehyde-derived adduct in light of the previous work on the formation of hydroxymethyl etheno dA *in vitro* and *in vivo* (8, 11-16). In *in vitro* primer extension assays, 8-HM-εC has been found to be a miscoding lesion when mammalian DNA polymerases were tested (19). In addition, the adduct itself is a structural analogue of 3,N^4-ethenocytosine (εC) (Figure 1), a well characterized exocyclic adduct (for review, see ref 53). It is of great interest for us to understand how the changes in adduct structure affect enzymatic recognition or catalytic efficiency (for review, see ref 54).

In this study, we first addressed the question whether the *E. coli* Mug protein, which excises εC (26,32,44), also acts on the newly synthesized 8-HM-εC. As shown in Figure 2, Mug is capable of binding to an 8-HM-εC-containing oligomer duplex in a protein concentration-dependent manner, similar to the binding of εC-duplex. Previous studies (32, 55-58) have revealed that both Mug and TDG bind strongly to their reaction product, *i.e*. an AP site-containing DNA when U•G- or T•G-containing substrates are used. In our binding experiments with the two exocyclic adducts, it is anticipated that the mechanism would be similar. Therefore, the binding of Mug to both adduct-containing oligomers in Figure 2 could

be part of a process involving the initial recognition and also should reflect the generation of an AP site. Nevertheless, the latter point concerning the AP site binding should serve as indirect evidence that Mug possesses a DNA glycosylase activity toward both adducts.

Using an oligomer cleavage assay, Mug protein showed a DNA glycosylase activity toward 8-HM-εC by forming an AP site at the adduct position. The rate of 8-HM-εC excision is ~2.5-fold lower than that of εC excision (Figure 3B). These two activities showed a similar pattern in their rate change during the course of their reaction to that observed with a G•U-containing duplex (32), *i.e*. the initial fast phase and a slow phase afterward. In the absence of *E. coli* EndoIV, Mug protein could not process more than its molar concentrations of the substrates (Figure 3B and C). This is due to the known fact that Mug strongly binds to the reaction product AP site-containing DNA, resulting in the lack of Mug turnover (32). Addition of EndoIV to the Mug reaction, a BER enzyme which cleaves the AP site next to a glycosylase action, enhanced the Mug activity toward 8-HM-εC by ~2-fold (Figure 3C). Moreover, the amount of 8-HM-εC removed was more than the molar concentration (1 nM) of Mug (reached more than 1.5 nM after 60 min), clearly indicating the turnover for the substrate. The same mechanism has been studied previously for both Mug and TDG using G•T(U)- and AP-site containing oligomer duplexes (32, 55,57), which revealed that EndoIV/HAP1 displaces the tightly bound Mug/TDG from the AP site-containing oligonucleotide. It should be pointed out that these studies showed a higher efficiency of turnover than did our results after addition of an AP endonuclease, which could be due to the fact that Mug bound to an adduct-containing oligomer may be more "difficult" to be displaced by an AP endonuclease than that bound to a normal base mismatch such as G•T(U).

The present finding that Mug also acts on 8-HM-εC can be reasoned that this adduct closely resembles the structure of εC. From the unrestrained molecular dynamics simulation,

the alignment and hydrogen bond pattern of 8-HM-εC•G pair is similar to that of εC•G pair (Figure 6). The latter structure agrees in principle with that reported using NMR as described in the Results section, which would validate our computational approach. Both adducts formed a highly sheared pair with the opposite G and have the same puckered sugar conformation. Such a structural motif observed in εC-containing duplexes was suggested to be important for its recognition by a specific DNA glycosylase (49).

The crystal structure of Mug complexed to an oligomer with a dU analogue•G mispair has been solved recently (59). From the same study using molecular modeling with εC, it is understandable why εC is a substrate for Mug. The authors showed that εC can fit into the non-specific pyrimidine pocket of the enzyme with its etheno moiety being comfortably accommodated in the hydrophobic space at the bottom of the pocket (59). In our case, the ability of Mug to recognize and remove 8-HM-εC indicates that this εC analogue should also fit into the active site. However, the lower Mug activity toward 8-HM-εC suggests some degree of steric hindrance to the binding and/or catalytic activity as a result of the hydroxymethyl group on the etheno ring. Similar examples can be found from the literature (for review, see ref 54). One example is from the study by Pegg and colleagues (60) in which the authors reported that the rates of removal of O^6-alkylguanines by *E. coli* O^6-alkylguanine-DNA alkyltransferase were affected by the size of the residue. The larger the size of the residue the slower the rate of the enzymatic activity toward the substrate (methyl > ethyl >hydroxyethyl).

For both 8-HM-εC and εC activities, like uracil excision, Mug protein requires double-stranded oligomer substrates. This enzyme differs from Ung protein in that Mug removes lesions only from duplex DNA (26,32) (also designated as double-stranded uracil-DNA glycosylase (27)), while Ung has about 2-fold preference for excising U from a U•G pair in single-stranded DNA compared to double-stranded DNA (61). When duplexes with modified

base mismatches were used, Mug showed a minimal degree of preference toward any 8-HM-εC or εC mispairs (Table 1). With the εC adduct, previous NMR studies have shown that it cannot form strong hydrogen bonds with any opposite base (47-50). In contrast, none of these modified base mispairs is recognized by Ung or mutY. Ung has basically no sequence homology with Mug but shares structural similarity (62). Even though both enzymes recognize a U•G mismatch, Ung's tighter binding pocket seems to prevent the enzyme from accommodating εC or 8-HM-εC. The adenine-specific mismatch glycosylase, MutY, also seems not to act on any of the above modified base mispairs, indicating a strict requirement for its active site interaction.

Of all the *E. coli* glycosylases tested in this work, Mug is presently the only known *E. coli* enzyme excising 8-HM-εC or εC. The latter activity has been previously shown to be missing in cell-free extracts from a *mug* mutant (44), indicating that Mug may be the only glycosylase acting on εC in *E. coli*. The εC activity of Mug protein is known to be conserved during evolution since the homologous human TDG also excises εC (25,26). At present, there seems to be no way to determine whether these etheno and related derivatives are the original selective pressures under which Mug protein evolved or whether they are just fortuitously able to fit into the active site of an evolved protein. Considering the high miscoding potential of εC as well as 8-HM-εC, it is reasonable to assume that cells possess a specific repair enzyme toward these adducts.

REFERENCES

1. Knaap, G. A. C., Voogd, C. E., and Kramers, P. G. N. (1982) *Mutat. Res. 101*, 199-208.

2. Van Duuren, B. L., Langseth, L., Orris, L., Baden, M., and Kuschner, M. (1967) *J. Natl. Cancer Inst. 39*, 1213-1216.

3. Van Duuren, B. L. (1969) *Ann. N.Y. Acad. Sci. 163*, 633-651.

4. International Agency for Research on Cancer (1976) in *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Man* (11), pp 175-181, IARC, Lyon, France.

5. International Agency for Research on Cancer (1987) in *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans* (Supplement 7), IARC, Lyon, France.

6. U.S. Department of Health and Human Services, NTP (1998) in *8th Report on Carcinogens* (1), III667-669.

7. Bentley, P., Bieri, F., Kuster, H., Muakkassah-Kelly, S., Sagelsdorff, P., Staubli, W., and Waechter, F. (1989) *Carcinogenesis 13*, 321-327.

8. Steiner, S., Hönger, G., and Sagelsdorff, P. (1992) *Carcinogenesis 13*, 969-972.

9. International Agency for Research on Cancer (1985) in *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans* (36) pp 181-188, IARC, Lyon, France.

10. International Agency for Research on Cancer (1989) in *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans* (47) pp 237-261, IARC, Lyon, France.

11. Goldschmidt, B. M., Blazej, T. P., and Van Duuren, B. L. (1968)*Tetrahedron Lett.* 1583-1586.

12. Golding, B. T., Kennedy, G., and Watson, W. P. (1990) *Carcinogenesis 11*, 865-868.

13. Steiner, S., Crane, A. E., and Watson, W. P. (1992) *Carcinogenesis 13*, 119-124.

14. Van Duuren, B. L., and Loewengart, G. (1977) *J. Biol. Chem. 252*, 5370-5371.

15. Golding, B. T., Slaich, P. K., and Watson, P. (1986) *J. Chem. Soc. Chem. Commun*. 515-517.

16. Golding, B. T., Slaich, P. K., Kennedy, G., Bleasdale, C., and Watson, W. P. (1996) *Chem. Res. Toxicol. 9*, 147-157.

17. Kohwi, Y. (1989) *Carcinogenesis 10*, 2035-2042.

18. Chenna, A., Perry, A., and Singer. B. (2000) *Chem. Res. Toxicol. 13*, 208-213.

19. Singer, B., Medina, M., Zhang, Y., Wang, Z., Guliaev, A. B., and Hang, B. *Biochemistry*, accepted.

20. Bartsch, H. (1999) in *Exocyclic DNA adducts in mutagenesis and carcinogenesis* (Singer, B. and Bartsch, H., Eds.) pp 1-16, IARC Scientific Publications (150), Lyon, France.

21. Barbin, A. (2000) *Mutat. Res. 462(2-3),* 55-69.

22. Dosanjh, M. K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., and Singer, B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1024-1028.

23. Hang, B., Chenna, A., Rao, S., and Singer, B. (1996) *Carcinogenesis 17*, 155-157.

24. Hang, B., Singer, B., Margison, G. P., and Elder, R. H. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 12869-12874.

25. Hang, B., Medina, M., Fraenkel-Conrat, H., and Singer, B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 13561-13566.

26. Saparbaev, M., and Laval, J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 8508-8513.

27. Gallinari, P., and Jiricny, J. (1996) *Nature 383*, 735-738.

28. Horst, J.P., and Fritz, H.J. (1996) *EMBO J. 15*, 5459-5469.

29. Neddermann, P., and Jiricny, J. (1993) *J. Biol. Chem. 268*, 21218-21224.

30. Neddermann., P., and Jiricny. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1642-1646.

31. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J. Wiebauer, K., and Jiricny, J. (1996) *J. Biol. Chem. 271*, 12767-12774.

32. Sung, J-S., and Mosbaugh, D. W. (2000) *Biochemistry 39*, 10224-10235.

33. Hang, B., Chenna, A., Sági, J., and Singer, B. (1998) *Carcinogenesis 19*, 1339-1343.

34. Dosanjh, M.K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., and Singer, B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1024-1028.

35. ABI 392 model DNA synthesizer manual, Appendix 1, pp1-6.

36. Rydberg, B., Qiu, Z.-H., Dosanjh, M. K., and Singer, B. (1992) *Cancer Res. 52*, 1377–1379.

37. Michaels, M. L., Cruz, C., Grollman, A. P., and Miller, J. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7022–7025.

38. Manuel, R.C., and Lloyd, R. S. (1997) *Biochemistry 36*, 11140-11152.

39. Williams, S.D., and David, S. S. (1998) *Nucleic Acids Res. 26*, 5123-5133.

40. Williams, S.D., and David, S. S. (2000) *Biochemistry 39*, 0098-10109.

41. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L. *et al*., (1997) *AMBER 5.* University of California, San Francisco.

42. Guliaev, A. B., Sagi, J., and Singer, B. (2000) *Carcinogenesis 21*, 1727-1736.

43. Lavery, R., and Sklenar, H. (1996) *CURVES 5.1. Helical analysis of irregular nucleic acids*. Paris: Laboratory for Theoretical Biochemistry, CNRS.

44. Lutsenko, E., and Bhagwat, A. S. (1999) *J. Biol. Chem. 274*, 31034-31038.

45. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560-564.

46. Sági, J., Hang, B., and Singer, B. (1999) *Chem. Res. Toxicol. 12*, 917-923.

47. Cullinan, D., Korobka, A., Grollman, A. P., Patel, D. J., Eisenberg, M. and de los Santos, C. (1996) *Biochemistry 35*, 13319-13327.

48. Cullinan, D., Johnson, F., Grollman, A. P., Eisenberg, M., and de los Santos, C. (1997) *Biochemistry 36*, 11933-11943.

49. Cullinan, D., Eisenberg. M., and de los Santos, C. (1999) in *Exocyclic DNA adducts in mutagenesis and carcinogenesis* (Singer, B. and Bartsch, H., Eds.) pp 179-189, IARC Scientific Publictaions (150), Lyon, France.

50. Korobka, A., Cullinan, D., Cosman, M., Grollman, A. P., Patel, D. J., Eisenberg, M., and de los Santos, C. (1996) *Biochemistry 35*, 13310-13318.

51. Cullinan, D., Johnson, F., and de los Santos, C. (2000) *J. Mol. Biol. 296*, 851-861.

52. Singer, B., and Grunberger, D. (1983) *Molecular Biology of Mutagens and Carcinogens*, pp 1-347, Plenum Press, New York.

53. Leonard, N. J. (1992) *Biochem. Mol. Biol. Chemtracts 3*, 273-297.

54. Singer, B., and Hang, B. (1997) *Chem. Res. Toxicol. 10*, 713-732.

55. Waters, T. R., and Swann, P. F. (1998) *J. Biol. Chem. 273*, 20007-20014.

56. Schärer, O. D., Nash, H. M., Jiricny, J., Laval, J., and Verdine, G. L. (1998) *J. Biol. Chem. 273*, 8592-8597.

57. Waters, T.R., Gallinari, P., Jiricny, J., and Swann, P. F. (1999) *J. Biol. Chem. 274*, 67-74.

58. Hardeland, U., Bentele, M., Jiricny, J., and Schär, P. (2000) *J. Biol. Chem. 275*, 33449-33456.

59. Barrett, T. E., Schärer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) *EMBO J. 18*, 6599-6609.

60. Pegg, A. E., Scicchitano, D. and Dolan, M. E. (1984) *Cancer Res. 44*, 3806-3811.

61. Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) *Biochemistry 34*, 6109-6119.

62. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) *Cell 92*, 117-129.

Table 1. Recognition of 8-HM-εC or εC mispairs by Mug protein[a]

Opposite base	8-HM-εC	εC
G	100[b]	100
A	62	111
T	74	143
C	95	137

[a] The standard reaction conditions were used for this experiment (See EXPERIMENTAL PROCEDURES). The incubation time was 10 min at 37°C. Data is an average of three independent experiments. No activity was detected using MutY or Ung for these mispairs.

[b] Relative excision activity. The initial rates of excision activity measured on 8-HM-εC or εC were treated as 100%.

Table 2. Intra-base pair parameters for the C•G, εC•G and 8-HM-εC•G basepairs

Basepair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
C•G	0.15	0	-0.02	8.12	-0.11	-0.86
8-HM-εC•G	1.38	1.0	-0.74	-5.99	-14.26	28.06
εC•G	2.01	0.7	0.1	5.11	-14.04	33.47

FIGURE LEGENDS

Figure 1. Chemical structures of 8-(Hydroxymethyl)-3,N^4-ethenodeoxycytidine (8-HM-εdC) and 3,N^4-ethenodeoxycytidine (εdC).

Figure 2. Binding of *E. coli* Mug protein to a 25-mer oligonucleotide duplex containing either 8-HM-εC or εC using a gel mobility shift assay. The reaction mixtures were incubated for 15 min at room temperature with increasing amounts of Mug protein. Lanes 4 to 9 and 11 to 16 contained 0.32, 0.8, 1.6, 3.2, 8.0 and 16 nM Mug protein. Lane 2 contained 16 nM Mug. Lanes 1, 3, and 10 contained buffer only. B: bound; F: free DNA.

Figure 3. *A*, Protein-dependent cleavage of 25-mer oligonucleotide containing either 8-HM-εC or εC by *E. coli* Mug protein. Increasing concentrations of Mug protein (0-4 nM) were incubated with ^{32}P-end labeled oligomer substrates (1.5 nM) for 1 h at 37^{o}C. Details see EXPERIMENTAL PROCEDURES; *B*, Time-dependent cleavage by *E. coli* Mug protein of ^{32}P-end labeled 25-mer oligonucleotide containing either 8-HM-εC or εC. The oligomer duplexes (2 nM) were reacted with 0.5 nM Mug protein for varying times at 37^{o}C; *C*, Stimulation of Mug activity toward 8-HM-εC by *E. coli* EndoIV. 4 nM of ^{32}P-end labeled 25-mer duplex was incubated with 1 nM Mug protein with or without 4 nM EndoIV at 37^{o}C. 1 mM $MgCl_2$ was used for these reactions. Reactions were stopped at various time points by heating at 95-100 ^{o}C for 3 min. 10 nM EndoIV was then added to all samples including those with Mug plus EndoIV before heating. After 20 min at 37^{o}C, reactions were terminated by adding a F/E solution.

Figure 4. Mug excision of 8-HM-εC with and without competitors. The standard reaction mixtures were incubated for 10 min at 37^{o}C with increasing amounts of unlabeled 25-mer oligomer duplex containing either 8-HM-εC (lanes 3 to 6), or εC (lanes 7 to 10) or

unmodified C (lanes 11 to 14). Lane 1 contained buffer only. Lanes 2 to 14 contained 4 nM Mug protein.

Figure 5. Screening of activities of *E. coli* DNA glycosylases toward oligomers containing either 8-HM-εC or εC. The substrate oligomer duplexes were incubated for 1 h at 37^{o}C with each individual purified DNA glycosylases: Mug: 0.01 U; Ung: 0.1 U; MutY: 0.1 U; Fpg protein: 0.1 U; EndoIII: 0.1 U and EndoVIII: 0.1 U. In addition, a cell-free extract prepared from a wild-type *E. coli* strain, BW 32, was also tested for activities against the modified bases (Lanes 8 and 16). Lanes 1 to 4 in both panels also contained 5 ng of HAP1.

Figure 6. Top view of the central 3 basepair motif for the control (C•G) and two lesion-containing DNA duplexes (εC•G and 8-HM-εC•G respectively) produced by molecular modeling. The central C•G, εC•G and 8-HM-εC•G basepair are colored by atom type. The hydrogen bond patterns for each base-pair are shown in dotted yellow lines. Note the similar displacement of the adducts toward the major groove and similar sugar conformations in the εC•G and 8-HM-εC•G basepairs.